Massachusetts Institute of Technology
Artificial Intelligence Laboratory
February 1976

AI WORKING PAPER #119 LOGO WORKING PAPER #47

A KNOWLEDGE-BASED COMPUTER ANIMATION SYSTEM

by

Kenneth M. Kahn

Abstract

This paper reproduces part of a doctoral thesis proposal describing the design of a system capable of generating animated drawings in response to a simple story. The representation and interaction of the various sources of the knowledge necessary to accomplish this are discussed. The appropriateness of an actor formalism for representing the concurrent processes and knowledge of the system is touched upon here and discussed further in Working Paper 120 (or Logo WP 48) "An Actor-Based Computer Animation Language". Finally, the role of the system as an example of a visible intelligent system in education is discussed.

This work was supported in part by the National Science Foundation under grant number GJ-1049 and conducted at the Artificial Intelligence Laboratory, a Massachusetts Institute of Technology research program. Reproduction of this document in whole or in part is permitted for any purpose of the United States Government.

TABLE OF CONTENTS A Knowledge-based Computer Animation System

Please send any comments to:

Kenneth Kahn
c\o MIT Logo Lab Room 348
545 Technology Square
Cambridge, Ma. 02139

Section I: The Problem

Computer animation is very limited compared with hand drawn animation. At the present, it is well adapted for producing abstract, typically geometrical, animation. To produce representational or narrative animation of people and trees and cars that move and look something like the real things is extremely time consuming even with the help of the computer. On the other hand, conventional animation, which is typically better quality, is very time consuming and requires expensive equipment. As such it is available as a medium of communication and expression to very few adults and even fewer children. Regardless of the type of animation, many skills are necessary to transform a story board into a finished cartoon.

I am proposing to attempt to alleviate this problem by equiping a computer with many of the skills necessary to go from an idea to a finished cartoon. This research diverges from the usual computer animation and graphics research by facing the problem of putting the required knowledge into a computer, rather than relying on the user. In order for the computer to be a useful apprentice in this endeavor it must know many animation and cartooning techniques such as in-betweening, easing in and out, anticipation, exaggeration, flow of motion, composition, and perspective. It must also know many things about the real world: if a person is described as happy then probably that person is smiling, if someone says "Hi" there is good chance that person raised his or her hand in greeting, people walk by moving their arms and legs in a particular pattern and so on. The computer needs to able to accept statements about a character or object and take those facts into account while constructing the animated scenes. Knowledge of the world, the characters, and the animation techniques must interact with the desired story in such a way that a reasonable approximation of the user's idea is converted into a cartoon.

Clearly, this is a huge problem. Therefore I plan to explore only a limited portion of this problem. The hope is that, if this is successful, a more complete system could follow. The major limitations envisioned are: [1]

(1) The only characters the system will know about are a few of Schulz's Peanuts characters,

(2) The computer will understand only simple activities. For example, it will understand walking, running, talking, turning and the like, but will *not* understand more complex or unusual activities such as fighting, skipping rope, and mountain climbing.

(3) The drawings will be simple line drawings,

(4) The English that the user uses to communicate with the system will need to be simple and stylized.

1 More limitations will probably become necessary as this research progresses.

Section II: Motivation for Creating such a System

The reasons for building such a system are many and varied.

(1) Artificial Intelligence (AI): This research is being carried out at the MIT Artificial Intelligence Laboratory where the problems of instilling intelligence in computers and of understanding natural intelligence are being explored. The animation system should be a good domain within which to explore the problems of representing real-world knowledge and common sense, and of handling interactions of many different sources of knowledge, such as animation techniques, cartooning techniques, the Peanuts characters and their world, human emotions and how they influence people's appearance, and the story presented by the user.

(2) Educational: The system could provide a different and stimulating environment for children to learn to express themselves well, to learn about animation and film in the role of a director, to learn how to criticize their own work in order for it to be improved, and to get experience working on relatively long-term projects. These activities fit well with the activities of the LOGO Laboratory,[2] a part of the Artificial Intelligence Laboratory, which among other things involves children in learning by programming within a specially designed computer language on well-chosen problems. The teaching of programming to children and this intelligent animation project can be more directly linked if the children are encouraged to modify

---

2 This research is being sponsored by the LOGO Lab. The original idea of this project came from Professor Ira Goldstein of the Lab and was initially explored in a seminar led by him.

and add new things to the animation system, in addition to simply telling stories and criticizing resulting cartoons. If the system is designed and implemented properly then there is a good chance that the children will be able to learn enough about how the system works to consider it a "white box."[3] That is they might understand how the system "thinks", thereby providing them with an example of thinking to help them in being more explicit about their own thinking.

(3) Computer Graphics and Animation: To my knowledge, this will be the first time that a computer animation system knew what it doing, in addition, to doing it. This knowledge should enable the system to make reasonable guesses, rather than require the user to explicitly describe each detail. For example, the system will, unless told otherwise, make assumptions as to whether a particular character is smiling, where to place the characters and background, how fast to move things and the like. In addition, the system is to be built upon primitives that are very well adapted to animation. They are based on the "actor" semantics of computation of Carl Hewitt ([Hewitt 1975] and [Smith 1975]) and Alan Kay [Goldberg 1974]. The relation of this research to this advanced programming language technology is discussed in the companion AI Working Paper 120 (or Logo 48).

(4) Entertainment: This may be the first time that animation would become an *active* event in people's activities, as opposed to the usual passive viewing of other's work. This project may be a pioneer in systems that, as opposed to television, would encourage the viewer to create besides viewing animation.

---

3 The term "white box" is meant to contrast with "black box" and is discussed in [Goldstein 1976].

Section III: The Basic Idea

The system, as presently envisioned, will go through the following steps in its normal mode of operation:

(1) translate the user's story or script into an internal representation;

(2) for each episode in the story, generate symbolic descriptions of successive key-frames;[4]

(3) find differences between the descriptions of successive key-frames, thereby generating a description of the transformation between them;

(4) use the transformation or difference description to generate commands to the appropriate "actors" which result in commands to the display;

(5) if the user criticizes the resulting cartoon, then alter the appropriate key-frames and transformations and repeat steps (3) through (5).

If at any point in the performance of steps 1 through 3 the system is confused or forced to make too many unfounded assumptions, then a dialog would be initiated with the user and advice would be sought.

---

4 By the term, "key-frame", I mean "film" frames that are interpolated between. This is not to be confused with Minksy frames which are closer to what I call "episodes" later.

Section IV: An Example

An imaginary scenario is appropriate here. Suppose the system was told the following story:

> Charlie Brown walks up to Lucy and greets her by saying, "Happy New Year, Lucy!" The background is a grass field.
>
> Lucy asks, "Does that make it so?" Charlie Brown has a look of puzzlement and interest. Lucy continues, "Does your saying, 'Happy New Year' MAKE it happy?" Lucy's arms are upraised as sign of emphasis. Lucy goes on, "Just because you SAY it does that mean it WILL be?" Lucy's arms are outstretched as sign of question. Charlie Brown has a facial expression of confusion.
>
> Lucy shouts, "IS THIS A GUARANTEE? IS THIS ..."
>
> Charlie Brown says, "Oh, good grief."

This may be translated into the following internal representation:[5]

```
(Background grass-field)
(key-frame 1
    (Charlie Brown (walking (towards Lucy))
                   (says "Happy New Year, Lucy!")
                   (greeting Lucy))
    (Lucy (on screen)))
(key-frame 2
    (Lucy (asks "Does that make it so?")))
(key-frame 3
    (Charlie Brown (expression (and interest puzzlement)))
    (Lucy (arms up-raised)
          (expression emphatic)
          (asks "Does ... ?")))
(key-frame 4
    (Charlie Brown (expression confusion))
```

---

5 Note that the details of this example are *ad hoc* since the system is at such an early stage of development.

```
          (Lucy (asks "Just because ...")
                (arms out-stretched)
                (expression questioning)))
     (key-frame 5
          (Lucy (shouts "IS THIS A GUARANTEE?")))
     (key-frame 6
          (Charlie Brown (says "Oh, good grief"))
          (Lucy (continues "IS THIS ...")))
```

The next task of the system is to fill into these key frames important details that are not explicitly given by the text. This is accomplished by the use of defaults and very simple common sense inferences. The "complete" descriptions of the key-frames follows (the notation "<<...>>" is used to distinguish the assumptions made):

```
     (key-frame 1
          (Charlie Brown (walking (towards Lucy))
               (says "Happy New Year, Lucy!")
               (greeting Lucy)
               <<arm extended>>
              ; based on the greeting sterotype
               <<expression mildly-friendly>>
              ;since a greeting is going on
               <<position screen-left>>
              ; this choice is arbitrary
               <<orientation full-right-profile>>
              ;based on the above position and "(walking (towards Lucy))"
               <<clothes default-Charlie-Brown>>
              ;a smarter version would know that "Happy New Year" indicates (clothes winter))
          (Lucy <<standing (facing Charlie Brown)>>
              ;This is assumed since she is talking with him
                 <<orientation full-left-profile>>
              ;since she is facing Charlie Brown
                 <<expression mild>>
              ;either default or for contrast with future frames
                 <<arms down>>;standard default
                 <<position screen-right>>
              ;since there are only 2 people and Charlie Brown already is "screen-left"
                 <<clothes default-lucy>>))


      (key-frame 2
          (Charlie Brown
```

```
                <<standing (facing Lucy)>> ;since he is talking with her
                <<arms down>>)
             ; people get tired of extending their arms, so if nothing to the contrary
             ; is indicated then they are lowered. The rest is "no-change"
        (Lucy (asks "Does that make it so?")))
             ; the rest is "no-change" and so need not be repeated
```

The other key-frames would also be expanded of course but will not be here.

Next the transformations are created based on the differences between the key frames. The transformation between the first two key-frames follows:

```
(Transformation key-frame1 key-frame2
        (Charlie Brown (walking standing)
                       (arms (one extended) down)
        ;note there is no position change except that implicit in "walking"
                       (talking listening))
        (Lucy (listening talking)))
```

This results in calls to the processes or actors involved. For example, the transformation "(Charlie Brown (arms (one extended) down))," (which means that Charlie Brown's arm changes from being extended in the first key-frame to where both arms are down) becomes a call to the "actor" Charlie Brown with the message "(arms down)". This actor or process interprets the message, making assumptions as necessary such as the use of his right arm. The movement would take into account reasonable speeds for such movements, easing in and out, and the like.[6]

After the "actors" are called, a movie consisting of a list of commands to the computer's display is produced and then shown. At that point, the user may enter criticism or advice. Some examples are:

"No, have Charlie Brown come in from screen right"
"Charlie Brown should walk in front of the tree not behind it"
"Ok, but the eyes in the last part should open wider."
"Lucy turned her head too fast."

- - - - - - - - - - - - - - - - - - -

6 This separation of phases, where static key-frames are created first and then differences and transformations are found, may be inappropriate. More examples need to be studied before this scheme can be evaluated.

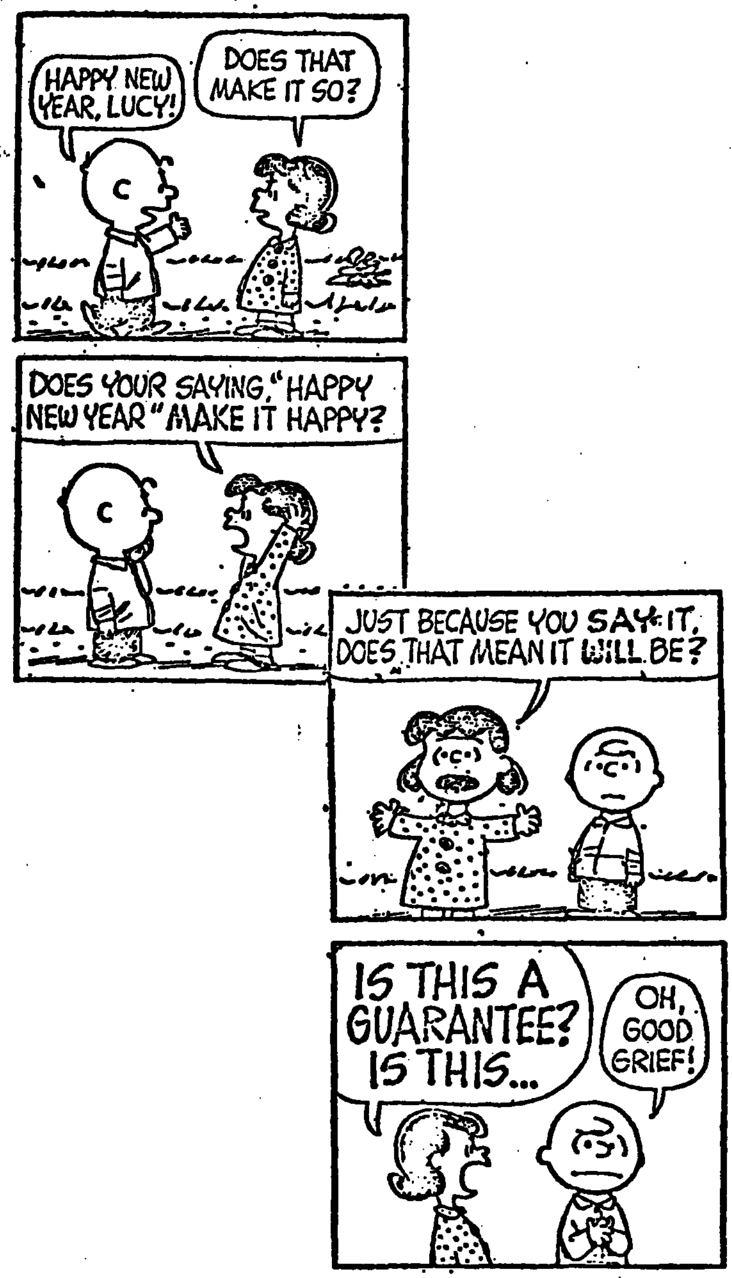
"We should zoom in on Lucy when she shouts"
"Half way through they should turn towards the 'audience'."
"Something is wrong in the last scene but I don't know what."
"Instead of the first cut try a fast dissolve."

It should be noted that understanding and determining what should be done when such comments are entered are very difficult problems.

The story in this scenario was based on a Schulz cartoon [Schulz 1962] which can be seen on the next page. The reader is invited to compare it with his or her image while reading this section.

Figure 1

Cartoon that Example is Based On



Section V: Key-Frames and Transformations

Key-frames are film frames that characterize a scene. They are symbolic descriptions of who and what is on the screen and in what form. They are essentially static descriptions of the world at a point in time. When enough things change a new key-frame is generated. For example, when the "camera" position changes significantly or when the characters change position, orientation, or expression significantly then a new key-frame describing the new situation is created. Of course, the new key-frame typically inherits much of its description from the previous frames.

Transformations are descriptions of the differences between two successive key-frames. Some elements will correspond to differences such as "before his arm was up and now it is at his side," while most will correspond to more dynamic descriptions of the change such as "his arm is being lowered until it comes to his side." Both of these examples indicate essentially the same action on the screen; however, there are cases where only one kind of description is appropriate. For example, "He ran very fast up to Lucy" is best represented as a dynamic transformation. On the other hand, if, for example, Lucy's expression is first described as angry and later as smiling, then a state change description would be more informative than describing Lucy as beginning to smile. This is because the action "beginning to smile" does not indicate the previous state which may influence the interpolation between the states. It should be noted that transformations are only partially derived from key-frame comparisons, sometimes they will be part of the initial story, for example, the story may have included the phrase, "he jumped up and down."

A) Semantic Interpolation or Symbolic Difference

Usually computer animation systems that can interpolate between two scenes do so without any knowledge of what the objects are that are being "in-betweened." Usually a point (or a line or a parameterized curve) in one scene is a associated with a point (line or curve) in the other. Then the creation of in-between positions for that point (or line) is computationally simple. There are at least two problems with this method though. First, points in the source are easily associated with the wrong target points. Second, natural motion it seldom produced. Take as an example a stick figure walking. A good speed and rate of easing in and out for the arms may not be the same for the legs. The rest of the body is also moving and yet not a constant rate. Meanwhile there may be a breeze going through a tree in the background that requires another rhythm.

The proposed knowledgable animation system would solve these problems in two ways. The interpolation is done not on the points or lines in the scene but on a *symbolic* description of the scene. This use of symbolic descriptions greatly simplifies the problem as has been demonstrated in many of the AI programs of the last five or ten years. Rather than comparing the points or lines of, say, Charlie Brown's face, we can compare simple descriptions of his eyes, mouth, ears, and face. The description of the differences between the symbolic descriptions of two key-frames is also more useful than a lower-level description in terms of points or lines. It may describe his mouth as first happy and then angry. In that case, the procedure that knows about mouths in general and the procedure that knows about any peculiarities of Charlie Brown's mouth would then be in a good position to generate the in-betweens for his mouth. The knowledge of how an effective change of expression appears is then brought to bear, rather than relying on some general interpolation mechanism. It should be noted that if several body parts interpolate in a similar manner then it should be easy for them to share the same knowledge.

B) The Vocabulary of Key-frames and Transformations

After it is decided that symbolic descriptions of scenes and transformations between scenes will be used, the question arises as to what an appropriate vocabulary for the objects, relations and actions is. To get a feel for the problems involved I studied the drawings of Charlie Brown's face in a couple of Schulz's Peanuts books. The outcome was encouraging for it seems as if almost any expression found in the books can be constructed selecting eyes, ears, mouths, etc. from a small "catalog" of parts of the head. For examples, see Figure 2 where they are listed pictorially.[7]

Relations between objects and actions have not been investigated very much yet. Hopefully, we will find that only a relatively small number of descriptors will suffice for the simple animation we have in mind. A small size vocabulary of actions and objects is desirable because knowledge must be attached to each descriptor about how to interpolate, how to make reasonable default judgements, and the like. One may object that by classifying objects, actions and relations into coarse groupings that subtleties that have important effects on the viewer are lost. This is a possibility, but it is premature to propose a computer AI system that would produce cartoons of anywhere near professional quality. Any lack of subtleties, while unfortunate, will not significantly affect the success of the system as a learning environment and as a prototype for subsequent more advanced systems. We must learn to walk before we attempt to run.

---

7 Mark Adler at the University of Edinburgh is doing vision research using Peanuts comics as a mini-world and may have interesting observations relevant to this cataloging task. [Adler 1975]

Figure 2 A

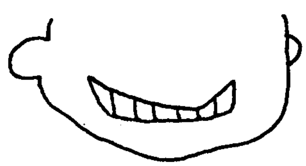
Vocabulary Sample -- Mouths

1. Short Frown

2. Short Dash

3. "Zig-zag" Smile

4. Teethy Smiling Profile



5. Teethy Smile

6. "Zig-zag" Frown

7. Very OPen Profile

8. Open Dark

Figure 2 B

Vocabulary Sample -- Eyes

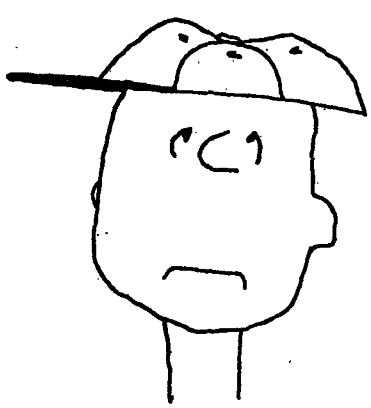
1. Dots with Short High Eyebrows

2. Vertical Dash (profile)

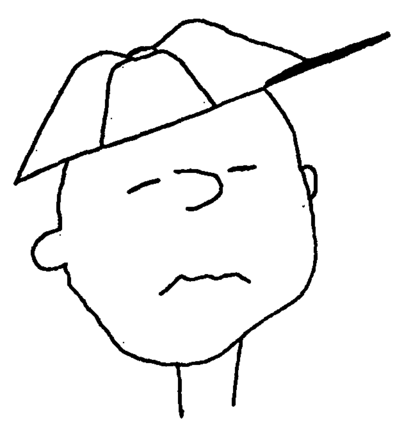
3. Ovals

4. Vertical Dash with Eyebrow

5. Vertical Dash with Eyebrows and "Surprise' line



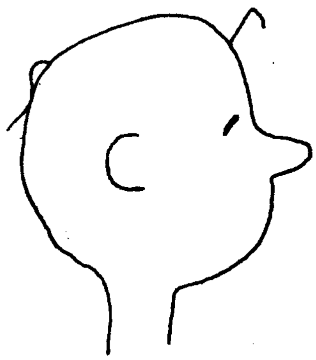
6. Two Ovals and Curves *(Looking up)



7. Horizontal Dashes

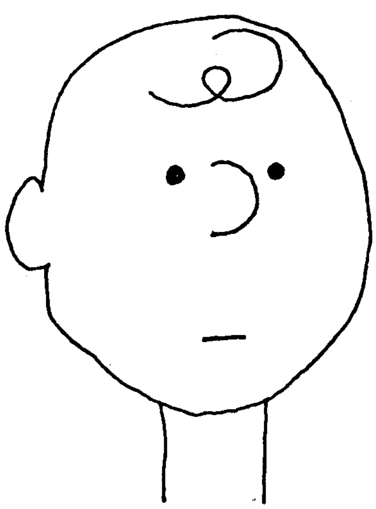
Figure 2 C

Vocabulary Sample -- Head Orientations

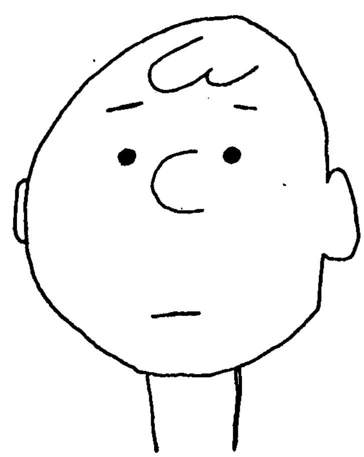


1. Right Profile

2. Left Profile



3. Right Front



4. Left Front

Note: There are surprisingly few intermediate views to be found.

Section VI: Defaults and the Role of the Episode

An episode is a coherent piece of the cartoon, usually separated from other episodes by a cut or dissolve. It is within each episode that key-frames are generated. In order for reasonable defaults to be chosen, the system needs to identify with each part of the story a sterotyped or generalized episode. These episodes are similar to what Schank calls "scripts" [Schank 1975] and Minksy calls "frames." [Minsky 1975]. For example, the system will have sterotypes for episodes such as, someone greeting another, or someone arguing with another, or people playing baseball. These structures contain much information, thereby lessening the demands on the user to describe the story in tedious detail. For example, once a story fragment is recognized as an instance of a greeting, then (unless there is reason to assume otherwise) the character will raise his or her hand while saying the greeting and perhaps the other character will nod. In addition, the sterotyped episode would contain such defaults as the normal duration of a greeting, the normal facial expressions and the like.

An episode provides much of the details necessary to produce the calls to the lower-level drawing and animation primitives. There are, however, other sources of defaults. Continuing with the greeting example, suppose the sterotype for greeting episodes indicates that Charlie Brown should raise his arm. The sterotype knows nothing special about Charlie Brown and so should not decide whether Charlie Brown should use his left hand or right hand. This knowledge would exist elsewhere, for example, in the description of a typical person which might indicate that if one hand is carrying something use the other, if both are busy then suggest that perhaps no hand should be raised, and if both hands are free then pick the right arm. Now it might be that

Charlie Brown is left-handed and so should act differently in the last case. This information would be associated with Charlie Brown and would be considered before more general information about people. In other words, I envision the use of a hierarchy of knowledge and appropriate inheritance operators.

Section VII: The Role of Animation Knowledge

It is obvious that an animator, in addition to knowing about the real world, knows much about animation techniques, many of which apply equally well to computer animation as to hand-drawn animation. For example, the animator knows how to ease into and out of a motion, how to exaggerate an action and how to produce a more dramatic effect by providing contrast to anticipated actions. A good animator also has a well-developed sense of timing and rhythm. He or she knows how long certain acts should take and how to mix different acts together to produce a rhythm in the film. Also, the animator knows how to produce film effects such as dissolves, zooms, pans and cuts. Probably most important an animator knows when to apply these techniques.

Clearly, I need to imbed these techniques and advice about their application into the computer animator. The first task I face is grouping these techniques into categories. For instance, dissolves, pans, zooms and cuts might all be grouped into "camera" techniques. Easing in and out and in-betweening perhaps should be grouped as "motion" techniques. Anticipation, exaggeration, and focusing attention perhaps should be called "dramatic" techniques.

Camera techniques could be procedures within the computer that are suggested by the episode sterotypes. Perhaps the greeting episode would suggest a pan from the person greeting towards the other at a medium "camera" distance. The argument episode may suggest that quick cutting and close-ups on the faces of the debaters would be best. Cuts would, in general, be used to separate episodes and so on. These would only be suggestions and a computerized "aesthetics" critic may notice that, say, too many dissolves have been going on and either veto the suggestion or propose a cut as an alternative.

The *motion techniques* could be procedures that are suggested by the object being affected. In that way trees would know how to react to wind in a way that is different from how an arm would move or an eye close. As is discussed later all these objects are imbedded in a hierarchy which describes what class each object is a member of, and in turn what class each class is a member of and so on. For example, Charlie Brown's left arm is an instance of Charlie Brown's arms, which are instances of arms (if adult arms were different, another class could be inserted), which are a kind of limb, which is a kind of body part, which is an object. Any advice as to the natural flow of an object is placed at the highest level of generality that is appropriate. If there is something special about arms that does not apply to legs then that information is associated with arms. In moving, say, Charlie Brown's left arm advice is searched for starting with the instance of the arm, then to Charlie's arms, then arms in general, and so on.

The problems of deciding when the *dramatic techniques* are appropriate is similar to the problem discussed in the next section on how the cartoonist's knowledge could be imbeded into the computer animator.

Section VIII: The Role of the Artist's or Cartoonist's Knowledge

An important aspect of animation is the drawings themselves considered as stills. The cartoonist's understanding of composition, balance, point of attention, contrast, and the like is important for producing aesthetically pleasing cartoons. These techniques along with the animator's dramatic techniques need to be incorporated into the system; however at this point, it is not too clear where. Some of the knowledge can go into the "aesthetics" critics mentioned earlier that reject or modify suggestions based on what preceded and on what will follow the current frames. In the process of filling in the key-frames this knowledge is used in a more positive way. Contrast and point of attention will be called upon to help place the characters and elements of the background whose position is under-specified. The episode will also have suggestions about the use of these techniques, for example, the argument episode knows that, in general, the end of an argument is the most dramatic part and perhaps should be emphasized by contrast, exaggeration and anticipation.

Section IX: Summary

This research while promising is at very early stages. The actor-based animation system, which is described further in the AI Working Paper 120 (or Logo 48), works but is in need of many improvements. The nature of key-frames, episodes, critics, and the vocabulary needed are being explored and much remains to be done. The educational aspect of the project is the most exciting part, yet little can be done now other than keep the design and representation simple and clear. I do hope to work with children as the system develops so that problems can be discovered early. Children will be using the actor-based animation system this spring, and hopefully improvements will become obvious and problems discovered at an early enough stage to do something about it. But as to whether the system will really work, and, if so, whether it will provide a new kind of environment to learn about animation, story telling, communication and thinking remains to be seen.

Bibliography

[Adler 1976]
Adler, M.
"Understanding Peanuts Cartoons"
Progress in Perception
Department of Artificial Intelligence Research Report No. 13
University of Edinburgh, December 1975

[Baecker 1969]
Baecker, R.
"Interactive Computer-Mediated Animation"
MIT EE Ph.D. Thesis 1969 MAC-TR-61

[Goldberg 1974]
Goldberg, A.
"Smalltalk and Kids -- Commentaries"
Learning Research Group, Xerox Palo Alto Research Center, 1974 Draft

[Goldstein 1974]
Goldstein, I. P.,
Understanding Simple Picture Programs,
MIT AI Laboratory AI-TR-294, September 1974

[Goldstein 1975]
Goldstein I., Lieberman H., Bochner H., Miller M.
"LLOGO: An Implementation of LOGO in LISP"
MIT-AI Memo 307, March 4, 1975

[Goldstein 1976]
Goldstein I., Papert S.
"AI, Language and the Study of Knowledge"
MIT-AI Memo 337, 1976

[Hewitt 1975]
Hewitt C., Smith B.
"Towards a Programming Apprentice"
IEEE Transactions on Software Engineering SE-1, March 1975

[Kahn 1976]
Kahn, K.
"Three Interactions Between AI and Education"
Machine Representations Of Knowledge
ed. Elcock, E., Michie D.

[Minsky 1975]
Minsky, M.
"A Framework for Representing Knowledge"
The Psychology of Computer Vision
ed. Winston, P.
McGraw-Hill Book Company, New York, 1975

[Papert 1971a]
Papert S.
"Teaching Children Thinking"
MIT-AI Memo 247, October 1971

[Papert 1971b]
Papert S.
"Teaching Children To Be Mathematicians vs. Teaching About Mathematics"
MIT-AI Memo 249

[Parke 1972]
Parke, F. I.
"Computer Generated Animation of Faces"
Univ. of Utah Report CSc-72-120, June 1972

[Schank 1975]
Schank, R.
Conceptual Information Processing
North Holland Publishing Company 1975

[Schulz 1962]
Schulz C.
"All This and Snoopy, Too"
Fawcett Crest Publications, Inc. Greenwich, Conn. 1962

[Schulz 1975]
Schulz C.
Peanuts Jubilee -- My Life and Art with Charlie Brown and Others
Holt, Rinehart and Winston, New York 1975

[Simmons 1975]
Simmons, R.
"The Clowns Microworld"
Theoretical Issues in Natural Language Processing
ed. Schank R. and Nash-Weber B. June 1975

[Smith 1975]
Smith B. and Hewitt C.

"A Plasma Primer"
MIT-AI Working Paper 92, October 1975

[Sutherland 1963]
Sutherland, I.
"Sketchpad: A Man-Machine Graphical Communication System"
MIT Lincoln Lab TR-296 1963